

## **The Effect of Income Inequality on Crime**

by Su Li, Caty McAfee, and Gaurav Phadke

### **Abstract**

Empirical research on the relationship between income inequality and of robbery and property theft. Our focus upon this literature lies within the visible manifestation of inequality with regards to the use of the Gini Index. Using the Gini Index as an indicator, we recorded an association between the distribution of wealth and crime. Our results attempt to link income inequality to the amount of violent crime by county. There seems to be evidence supporting our original hypothesis of income inequality contributing to the amount of violent crime, suggesting that such a disparity may be linked to social theories that contribute to the economics of crime.

## **I. Introduction**

Crime participation and its social consequences have long been recognized as significant within the field of economics. A long discussed matter, it can be traced back to the literature and philosophies of the great Ancient Greeks, including the works of Plato and Aristotle, and has been touched upon in many eras since. In recent times, studies that have focused upon the relationship between income levels and the rate of crime being the motivating reason for individuals to participate in crimes have come to light. However, this approach fails to consider other possible factors within its calculations, especially with regards the external features. We focus on the impact of the disparity of income within an area to be a critical factor in regards to the crime rate within the community.

Crime has long been associated with a variety of factors, particularly those on income levels, poverty levels, and exposure to a lack of human capital directly available to individuals. However, crime participation and the choice to be involved in such activities is linked through a cost-benefit calculation at an individual level for each criminal. The expected benefits of crime must be valued at a level that must outweigh the possible cost for the consequences of an illegal action. Alternatives to crime with a higher cost benefit ratio suggest that with an equivalent or higher payoffs, crime participation is predicted to fall, with the opposite also being true. With regards to income inequality within an area, benefits and costs are determined to have correlation with the Gini Index of a community. Higher disparities both increase and decrease the benefits possibly gained from a criminal venture while also affecting the baseline cost of one.

We constructed this research with the examination of the relationship between levels of crime within a county in the United States and the levels of income inequality within said county. Our focus is upon the Gini Index, a tool used to measure the disparity of income levels within an area, and its relationship to visible crime behavior within counties in the United States over a period of 5 years. Current economic theory on criminal behavior is largely influenced by the work of Becker (1968). He suggests that criminal activity is based on rational cost-benefit analysis by criminals and then by choosing the most profitable venture. One of the essential components to his theory is the size of the gap between the rich and the poor. The poor, who may be considered as the possible criminals, must size up the rich, the potential victims of crime. Especially in areas with significant disparities, criminal participation may be high when considering that the opportunity costs of such actions in areas of large

inequalities are relatively low in comparison to the potential and expected benefits of participation in crime. Thus, a high degree of inequality would lead to a higher risk of criminal participation.

## **II. Literature Review**

Hicks and Hicks (2014) explore conspicuous consumption rather than just simple income inequality in this paper. Traditional economic theory supports that greater degree of income inequality will correlate with a greater degree of crime. In this study, it was found that instead there is a direct correlation between conspicuous consumption and violent crime. That is to say, criminals were motivated by the inequality they saw from actual purchases rather than the actual total inequality. This motivated Hicks and Hicks (2014) to explore the possibility that information was a large part of crime as well.

Our paper is different firstly because we aren't basing our assumptions off conspicuous consumption, just total income inequality. Secondly, we are using different data. Hicks and Hicks (2014) got income information from the Consumer Expenditure Survey which usually measure only about 20,000 people a year whereas we got our results from the US Census which is more widespread. The CEX survey is also designed to be used at the state level whereas our data is by counties which gives more widespread results.

This paper focuses on the issue of crime and its approach to economic incentives. Much like legal workings of economic incentives, the criminal aspect must weigh the pros and cons of undertaking criminal and illegal ventures. The focus being upon two aspects of the findings, the relationship between criminal and legal opportunities, and the economic returns of either. According the paper's findings, both of these aspects are central points in the new and flourishing works of economics of crime as a research field. The findings of the paper show there are a variety of economic incentives that matter to the results of crime outcome.

Dahlberg and Gustavsson (2008) explore the difference between permanent income and transitory income and its affect on crime. Permanent changes in income can affect income inequality on both short and long term while transitory income only affects the short. They found that there was no significant correlation between an increase in transitory income inequality and crime. However, there was a great positive correlation between an increase in permanent inequality and crime.

This paper uses data from Sweden, which will obviously differ from ours as we are using United States statistics. These numbers also only cover 3.35% of the population. These numbers are county statistics, similar to ours, but we have a larger sets of data points with which to get a broader scope of the correlation.

Considered by many to be the founder of the field of the economics of crime, Becker lays (1968) the foundation to this field. In the paper, he considers criminals and their views towards modern interpretations of law and their costs to benefit ratio of conviction versus criminal activities. He puts forward many interesting points of view such as the balance of enforcement and punishment along with the costs of such activities. He argues that it is not completely possible or efficient to persecute all criminals and instead suggests that there is a combination of resources and punishments that minimize possible criminal participation and the social loss accompanying it. The optimal amount of enforcements depends upon a variety of variables in which he discusses in his paper.

### **III. Data**

#### **Data Sources and Description:**

The variables apart from the Gini coefficient were all chosen as indicators of changing incentives for the committing of crimes. In several studies, money for drugs was found to be one of the most commonly cited reasons for committing a property crime or robbery. Drug death data was added as an explanatory variable as a roundabout way of measuring the level of drug addiction in a county population. Drugs that are not as addictive/dangerous would not compel users to crime as much as drugs that were. Higher levels of drug-deaths in a county were hypothesized to correspond with higher crime rates. Unemployment rate was added as an explanatory variable as a measure of the level of incentive to commit a crime. If unemployment rate was low, obtaining money through obtaining a job would be much easier and would then disincentivize crime. The number of police officers in a county was added as an explanatory variable as it was hypothesized that a high numbers of officers would disincentivize crime by increasing the probability of being caught. This variable only measures active police officers and excludes civilians working for the department as well as office staff. Poverty level was added as an explanatory variable after the completion of our initial model in an attempt to better explain the Gini Coefficient's effect on crime. It was hypothesized that a high level of income inequality would likely not have a large positive impact on crime, if all people had incomes above the poverty line. On the other hand, a county with low income inequality, but with a large percentage of people living below the

poverty line, would have a higher rate of crime than the county in the first example. We hypothesized that a combination of poverty level and income inequality would better explain crime than either variable alone. The variable would compound with high values and negate with opposite values.

The data for the Gini coefficients by county as well as county poverty levels were obtained from the the U.S Census Bureau's American Community Survey results for 2014. The data for the robbery and property crime totals by county as well as number of officers employed were retrieved from the Uniform Crime Reports provided by the FBI. The data for county unemployment rates come from the Bureau of Labor Statistics.

### Defined Variables

*Table 1: Variables*

Label	Variable Description	Measure	Type
Crime	The number of crime	The number of total property crimes and robberies per hundred thousand people in counties in the United States in 2014	dependent
Gini	Gini coefficient	Income inequality across United States counties in 2014	independent
Death	Drug overdose death	The number of drug overdose death per county in 2014	independent
Unem	Unemployment rate	% of total workforce who are unemployed and are looking for a paid job in 2014	independent
Police	Employed Police Officers	The number of police officers employed per county in 2014	independent
Pov	Poverty Level	% of people in the county whose annual income between 2013 and 2014 was below the poverty line	independent

Table 2: Descriptive Statistics

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
crime	213	794.2	697.64	69.58	3784.75
gini	213	0.4462	0.0353	0.3301	0.5382
drug	197	13.05	3.46	5.05	19.05
ue_rate	213	6.58	2.44	2.9	23.6
policeOfficer	182	216.97	324.31	11.0	2674.0

#### IV. Initial Model

##### Description:

Our initial model looked at the relationship between Crime rate per 100,000 people vs the Gini coefficients for 213 counties in the United States. Our initial hypothesis was that increases in the Gini coefficient would correspond with an increase in the crime rate.

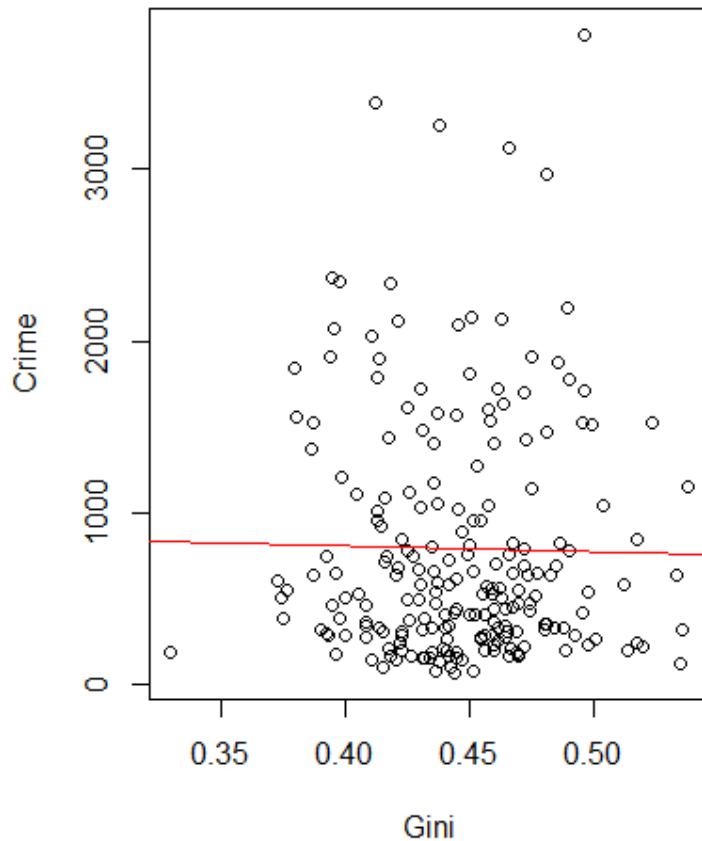
##### Initial Model Gauss Markov Assumptions:

1. It is assumed that our model can be written as  $\text{Crime} = \beta_0 + \beta_1(\text{Gini}) + u$ , meaning it is linear in parameters.
2. The data included every county in the United States that had complete and accurate data for Gini coefficients, crime rates, and population.
3. The values obtained for the Gini Coefficients are not all the same.
4. To test whether the error term has an expected value of zero and constant variance for all values of the explanatory variable, residuals were plotted against the corresponding Gini values. Based on the plot(see appendix), although the mean seems to around 0, it is likely that there is a non-constant variance.

##### Initial Model Results:

The equation from the initial simple regression model was found to be  $\text{Crime} = 950 - 349.2(\text{Gini})$ . A plot with the fitted line can be seen below. This result goes against our

initial hypothesis that higher values of the Gini coefficient would correspond to higher crime rates. This model is instead suggesting that every increase of .1 in the Gini coefficient would lead to a decrease of 34.92 crimes per 100,000 people. Although the observed model goes against our initial hypothesis, the p-value of the coefficient of the Gini variable was measured at .798, meaning that it is not statistically significant from 0 for any reasonable significance level. An R-value of .0003 indicates that the overall model explains almost none of the differences in crime rate. A complete overview of the initial model statistics can be seen in the expanded regression table below.



## V. Expanded Model

### Description:

In our expanded model, the variables drug overdose deaths per 100,000 people, unemployment rate, number of police officers, and poverty level were added. In addition the interaction effect between poverty level and Gini coefficient was added in an attempt to explain the non-relationship observed in the initial model. We hypothesized that high levels of income inequality would not have a significant effect on crime if incomes were high enough to ensure basic needs. The interaction between poverty

level and gini coefficient would help capture whether the effect of high levels of income inequality could be negated or exacerbated by different levels of poverty.

#### Expanded Model Gauss Markov Assumptions:

1. We assume the model in the population can be written as  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots\beta_kx_k + u$
2. The data for all additional variables was obtained as thoroughly as possible for all data points selected in the initial model.
3. None of the values of the independent variables are constant.
4. As stated in the initial model assumptions, the variance of the residuals does not appear to be constant when plotted against Gini values. This is also true for the interaction between Gini and Poverty, as well as Police data.
5. Based on the residual plots used in the previous assumption, it is likely that the error term is not normally distributed and independent of the explanatory variables.
6. Multicollinearity: although not an assumption, multicollinearity can have a negative effect on a multiple regression model. An analysis of the variance inflation factors found little to no multicollinearity between explanatory variables.

#### Expanded Model Results:

The table below shows a summary of various models tested. Coefficient values are listed first with t-values in parentheses. A description of significance is given below the table. All variables were added in the second model and then chosen variables were removed based on the observed significance levels.

Dependent Variable: Crimes Per 100,000 People					
Independent Variables	Model (1)	Model (2)	Model (3)	Model(4)	Model (5)
Gini Coefficient	-349.2 (-0.257)	-3670.53 (-0.893)	-3217.92 (-0.788)		
Drug Deaths per 100,000		9.298 (.626)			
Unemployment Rate		.3186 (.014)	-3.735 (-0.171)	5.459 (.250)	13.0749 (0.693)



Police Officers		.9167*** (5.676)	.9909*** (6.383)	.8837*** (5.822)	0.8736*** (5.803)
Poverty Level		51.9581 (.460)	60.198 (0.537)	7.1533 (.690)	
Gini * Poverty Level		-66.253 (-0.267)	-81.357 (-0.329)		
Intercept	950.0 (1.561)	1782.40 (984)	1701.87 (.943)	480.21*** (2.887)	542.97*** (3.906)
Degrees of Freedom	210	161	175	177	178
R-square	0.000314	0.1782	0.1951	0.1626	0.1603

**Signif. codes:**

0'\*\*\*' 0.01 '\*\*\*' 0.05 '\*\*' 0.1 '.' '1

### Robustness

Although Model 3 was found to have the highest R-value, it was compared to Models 4 and 5 to evaluate whether the addition of the extra variables resulted in a significant increase in overall significance. Setting the unrestricted model as model 3 and the restricted model as model 4, the F-statistic was found to be 3.541 with the area to the right of the F-statistic having a value of .0317. This indicates that the Gini coefficient, and interaction between Gini coefficient and poverty level are jointly statistically significant at the 5% level. When model 3 was compared to model 5, the F-statistic was found to be 2.5241, with the area to the right of the F-statistics having a value of .05927. This indicates that the Gini coefficient, poverty level, and interaction between them are jointly statistically significant at the 1-% level.

## VI. Analysis and Conclusion

Given the results of the Robustness tests, Model 3 was found to be the most appropriate. However, this model continued to have the problem of counter-intuitive signs on all of the independent variables except for poverty level. The sign for the Gini coefficient continues to be negative representing an inverse relationship between crime and income inequality. This goes against not only our initial hypothesis but other studies across countries indicating the positive relationship between the two. The interaction effect between poverty and gini coefficient was also found to be inverse in spite of the interaction being added in an attempt to alleviate the non-relationship between crime and the gini

coefficient. Unemployment and crime were found to have a negative relationship when it was thought that higher unemployment would indicate a larger proportion of the population who would have much more to gain from committing a robbery or property crime. Although the police variable went against our initial hypothesis, a possible explanation for the positive relationship is that it does not reflect a cause and effect relationship with crime, in that it is a lagged response to crime and not the deterrent to crime that we believed.

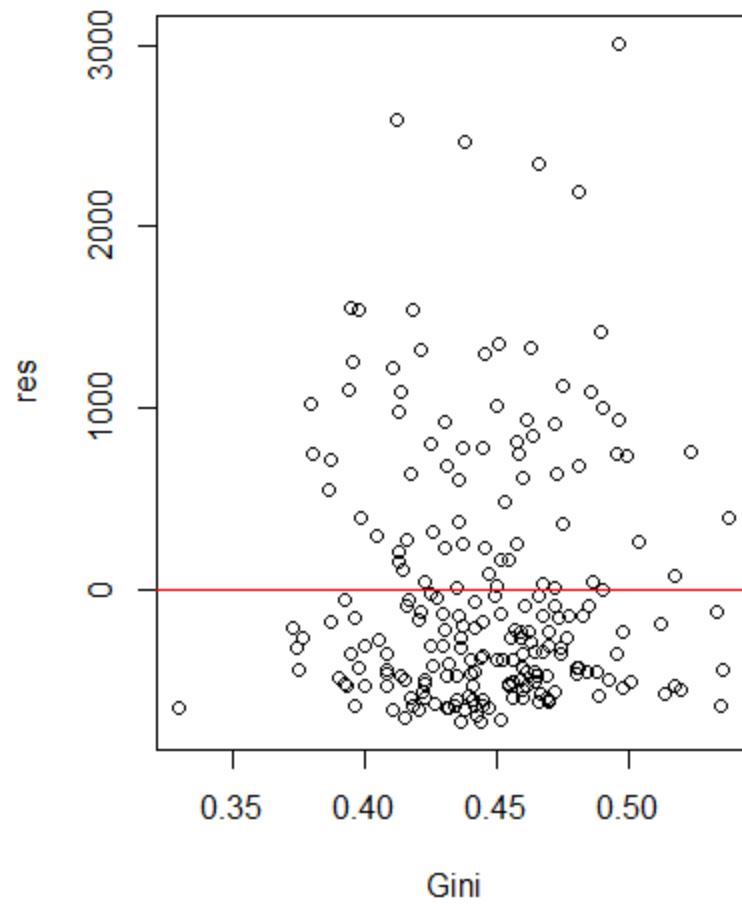
Possible explanations for the counter-intuitiveness of our model may include missing variables and incomplete data. A common problem with our data was the misrepresentation of crime rates in counties that contained large cities. It is not clear how the crime reports are created within a county but many appear to have been created without input from the city police departments. Obvious examples of this problem were deleted from our data-set and include counties containing cities such as Chicago, Oakland, San Jose, Los Angeles, and Phoenix. Although these data points were deleted, it is unclear how many other data points used had similar problems. Another likely problem is the lack of account for different cost of living across the United States. Poverty level does not vary geographically and so can be inaccurate in comparisons between counties with vastly different costs of living. Future studies should seek to use more locally accurate datasets as well as data that can accurately capture important differences between counties.

### References

- Mirko, Draca; Machin, Stephen (2015). "Crime and Economic Incentives." *Economic*. Vol. 7: 389-408.
- Hicks, Daniel; Hicks, Joan (2014). "Jealous of the Joneses: Conspicuous Consumption, Inequality, and Crime." *Oxford Economic Paper*.
- Dahlberg, Matz; Gustavsson, Magnus (2008). "Inequality and Crime: Separating the Effects of Permanent and Transitory Income." *Oxford Bulletin of Economics and Statistics*. Volume 70, Issue 2: 129-153.
- Becker, Gary (1968). "Crime and Punishment: An Economic Approach." *National Bureau of Economic Research*. P. 1-54.

## Appendix

Initial Model Residual Plot



## Model 1 Output

Call:

```
lm(formula = Crime ~ Gini)
```

Residuals:

Min	1Q	Median	3Q	Max
-725.4	-495.6	-247.2	270.9	3008.1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	950.0	608.7	1.561	0.120
Gini	-349.2	1359.8	-0.257	0.798

Residual standard error: 699.2 on 210 degrees of freedom

Multiple R-squared: 0.0003139, Adjusted R-squared: -0.004446

F-statistic: 0.06595 on 1 and 210 DF, p-value: 0.7976

## Model 2 Output

Call:

lm(formula = Crime ~ Gini + Death + Unem + Police + Pov + GiniPov)

Residuals:

Min	1Q	Median	3Q	Max
-1206.4	-424.7	-177.7	270.2	2570.1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1782.4024	1811.8203	0.984	0.327
Gini	-3670.5362	4110.3484	-0.893	0.373
Death	9.2989	14.8657	0.626	0.533
Unem	0.3186	22.0958	0.014	0.989
Police	0.9167	0.1615	5.676	6.27e-08 ***
Pov	51.9581	112.8688	0.460	0.646
GiniPov	-66.2537	248.5584	-0.267	0.790

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 646.6 on 161 degrees of freedom

(44 observations deleted due to missingness)

Multiple R-squared: 0.1782, Adjusted R-squared: 0.1476

F-statistic: 5.818 on 6 and 161 DF, p-value: 1.646e-05

### Model 3 Output

Call:

```
lm(formula = Crime ~ Gini + Unem + Police + Pov + GiniPov)
```

Residuals:

Min	1Q	Median	3Q	Max
-1474.6	-433.1	-191.8	252.2	2440.7

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1701.8712	1804.0396	0.943	0.347
Gini	-3217.9271	4084.6673	-0.788	0.432
Unem	-3.7351	21.8997	-0.171	0.865
Police	0.9909	0.1553	6.383	1.51e-09 ***
Pov	60.1980	112.1976	0.537	0.592
GiniPov	-81.3573	247.4051	-0.329	0.743

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 650.1 on 175 degrees of freedom

(31 observations deleted due to missingness)

Multiple R-squared: 0.1951, Adjusted R-squared: 0.1721

F-statistic: 8.486 on 5 and 175 DF, p-value: 3.271e-07

#### Model 4 Output

Call:

lm(formula = Crime ~ Unem + Police + Pov)

Residuals:

Min	1Q	Median	3Q	Max
-1501.8	-430.7	-187.3	308.7	2362.7

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	480.2091	166.3079	2.887	0.00437 **
Unem	5.4590	21.8732	0.250	0.80320
Police	0.8837	0.1518	5.822	2.68e-08 ***
Pov	7.1533	10.3680	0.690	0.49114

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 659.3 on 177 degrees of freedom

(31 observations deleted due to missingness)

Multiple R-squared: 0.1626, Adjusted R-squared: 0.1484

F-statistic: 11.45 on 3 and 177 DF, p-value: 6.706e-07



## Model 5 Output

Call:

```
lm(formula = Crime ~ Unem + Police)
```

Residuals:

Min	1Q	Median	3Q	Max
-1456.0	-426.6	-178.2	316.8	2388.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	542.9712	139.0186	3.906	0.000133	***
Unem	13.0749	18.8558	0.693	0.488954	
Police	0.8783	0.1514	5.803	2.93e-08	***

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 658.4 on 178 degrees of freedom

(31 observations deleted due to missingness)

Multiple R-squared: 0.1603, Adjusted R-squared: 0.1509

F-statistic: 16.99 on 2 and 178 DF, p-value: 1.764e-07

### Model 3 and 4 F-Test Output

#### Analysis of Variance Table

Model 1: Crime ~ Unem + Police + Pov

Model 2: Crime ~ Gini + Unem + Police + Pov + GiniPov

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	177	76944034				
2	175	73951060	2	2992975	3.5413	0.03107 *

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Model 3 and 5 F-Test Output

#### Analysis of Variance Table

Model 1: Crime ~ Unem + Police

Model 2: Crime ~ Gini + Unem + Police + Pov + GiniPov

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	178	77150964				
2	175	73951060	3	3199905	2.5241	0.05927 .

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1